

科技文献组织与揭示的基本框架

贾明静 王青

(中国科学技术信息研究所, 北京 100038)

摘要: 本文基于FRBR概念模型的思想, 提出一种基于MARC元数据揭示方法向资源描述与检索(RDA)转换的元数据框架, 打破依照文献类型逐章描述各自著录规则的体系, 构建统一通用基本框架, 从基本框架提取所需揭示的基本信息形成著录单元细分的元素集, 使得不同载体不同文献类型的著录框架具有高度的连贯性和一致性。本框架还可以继续沿用现有系统和编目人员, 通过对框架的设计机制让系统处理关联关系, 为将来选择RDA做好准备, 奠定基础。

关键词: 资源组织; 资源揭示; 元数据框架; FRBR; MARC

中图分类号: G254.364

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2014.06.010

Research on Organization and Reveal Basic Framework of Technology Literature Resources

Jia Mingjing, Wang Qing

(Institute of Scientific and Technical Information of China, Beijing 100038)

Abstract: Article give a metadata framework that is the conversion from MARC to RDA (Resource Description and Access) base on the conceptual framework of FRBR, break the cataloging rules system that in accordance with type of literature description, build unified, general and basic framework; and extract the basic information that need revealed from it, form elements set that minimize the granularity, make bibliographic framework of different carriers and different types of literature highly continuity and uniformity. From the operational point, this framework continue to use the system and cataloging staffs, the existing system is able to deal with the association based on the design mechanisms, this framework is foundation for select the RDA in future.

Keyword: resource organization, resource reveal, metadata framework, FRBR, MARC

1 引言

目前, 科技文献组织与揭示的方法主要有两大类, 即着重于资源描述的揭示方法和着重于资源关联的揭示方法。着重于资源描述的揭示方法是当前信息服务机构广泛使用的资源揭示方法,

包括基于AACR2编目体系的MARC元数据揭示方法与DC元数据揭示方法。MARC编目规则在国内科技资源描述上已有多年的应用, 而DC元数据揭示方法则是20世纪90年代开始兴起。成都理工大学的辜寄蓉博士等将成都各机构的信息采集起来转化为统一的元数据格式并建立基

作者简介: 贾明静* (1985-), 女, 中国科学技术信息研究所馆员, 研究方向: 信息组织; 王青 (1988-), 女, 中国科学技术信息研究所助理馆员, 研究方向: 信息组织。

收稿日期: 2014年8月13日。

于元数据的共享平台，实现了资源的聚合^[1]；浙江大学信息资源管理研究所的金更达等探讨了数字档案的集成，该文件集成方法集成了外部数据与内部资源，并将这些资源转化为元数据进行操作^[2]。着重于资源关联的揭示方法包括基于FRBR/FRAD框架的资源揭示方法与基于RDA（资源描述与检索）编目规则的揭示方法。国外关于FRBR的研究已经较为成熟，美国研究图书馆集团（RLG）的Red-Light-Green（红绿灯），是最早实现FRBR化的大型Web数据库，该机构将FRBR的“作品-内容表达-载体表现-单件”4级实体简化为二级“作品-版本（载体表现）”，并将编目记录全部上网，在其搜索结果显示页面显示的是作品，而不是一般OPAC的“载体表现”。Austlit^[3]网关是最早全面落实FRBR的数据库，专门适用于澳大利亚文本的文学性文集，该数据库以作品为中心，在一个Web页上为每个作品展示其所有的内容表达和载体表现。另外，Timothy J. Dickey将FRBR的概念应用于关联性较强的音像制品上，提供了音像资料的FRBR解决方案^[4]。国内在此方面的研究主要集中于FRBR概念模型，构建于FRBR概念模型之上的应用研究大多为建议和设想，还未有实际成果出现。谢琴芳在《FRBR用户任务对未来图书馆目录影响》中分析了FRBR用户任务与书目记录的关系，重点探讨了CALIS联合目录现状对用户任务实现的障碍及策略^[5]；鲍国强从实际工作出发，阐述古旧地图编目及数字化信息检索中应用FRBR基本模式的问题思考和解决之道^[6]。

RDA作为国际性的资源著录和检索的内容标准，引起了各大图书馆的密切关注^[7-8]。其中，澳大利亚国家图书馆还将实施RDA作为一项重要的战略任务正式写入了2009-2011年的战略规划中；美国斯坦福大学图书馆已经直接采用RDA对所有原始文献编目^[9]。相较于国外对RDA的研究已经进入实施阶段，国内学者对RDA多采取观望态度，开展了较多综述性研究。上海图书馆的刘炜与华东师范大学图书馆的胡小菁分别将RDA与关联数据和ISBD进行对比研究，探讨两

种元数据方案的异同之处。

上述两种资源描述方法有着根本的区别，前者对资源的揭示成扁平状结构，记录间较为孤立，同时容易产生数据冗余，在数据的规范上也并不完善；后者对资源的揭示呈现层级结构，各层级内部与层级之间关联性较强，弥补了前者的弊端，但尚处于探索阶段。

目前，大多信息服务机构在采购与编目环节已将不同类别的科技文献分开管理。这种管理模式及国内现行的AACR2编目体系直接导致部分科技文献的信息不能全部揭示。与此同时，国内广泛使用的科技文献著录或描述规则对科技文献自身的揭示较为孤立，这种孤立不仅表现在不同类型资源的揭示上，还表现在资源载体表现及内容表达的揭示中。此外，对科技文献间关系的揭示及关注也较少，也没有规定明确不同资源间的关系类型及揭示方法。

本文则以现有的编目框架为基础，基于FRBR/FRAD框架的关联揭示思想，构建了一个新的科技文献组织与揭示框架，以期解决编目人员对交叉性资源把握不准的问题，同时将庞大的文献集合勾画成一个包含各种载体、充分揭示各种关系、面向所有用户的一个结构化体系，为实现查找、识别、选择、获取的用户任务提供可能。

科技文献范围广泛，其中专利、标准等未采用MARC元数据编目，故本文所探讨的是采用MARC元数据编目的科技文献，如图书、会议录、期刊、科技报告等文献类型。

2 框架体系

在FRBR实体-关系模型构建了3组关系，其中与传统编目相关度较高的是第一组，即智力、艺术创作的产品。该组将书目资源分为作品-内容表达-载体表现-单件4个层级，如图1所示。该概念框架将作品的内容表达与载体表现与作品本身剥离开来专门揭示，以单件做为资源揭示的最小颗粒度。具体来说，作品的载体表现体现在载体的不同上，比如作品的文献类型可

能是图书、会议文献、期刊，载体介质可能是CD、DVD或者其他形式等；作品的内容表达体现在其表现形式的多样化，比如电影、音乐或其他形式。

在AACR2编目规则中，科技文献资源的揭示按文献类型的不同规定不同的著录规则，即以作品的载体表现层级为著录的最小单元，不利于对书目数据间层级关系的揭示。科技文献资源即为FRBR模型中的作品，它具有不同的表达方式和表现形式。因此基于FRBR概念框架，本文以单件作品为最小揭示单元，将科技文献资源分为包括责任者及主题等不同实体的规范记录构成的作品层、记录多种内容表达形式和载体表现形式的书目记录层以及由特定物理实体构成的馆藏层，并将该模型命名为WBC模型。另外，基于FRBR 3组实体间关系，即资源内部的关系、资

源与责任者的关系以及主题与资源间的关系，作品层以责任者（个人、团体）、作品及相关属性（资源本身）、主题（概念）作为3条主线对书目记录进行规范聚集，形成由规范实体构成的作品层。科技文献资源组织基本框架如图2所示。

从图2可以看出，每个层级间互相关联，同时层级内部也具有一定关联性。在作品层中，以作品及相关属性为核心规范记录，与责任者、主题相关联。规范记录与书目记录、馆藏记录相结合，通过相互之间的关联机制来继承属性、关系，并区分不同的属性和关系。在各层级内部和层级之间存在着4个关系主线：(1)整个模型的关系主线是以作品自身为主体，通过特定题名代表的作品关联到该作品的所有载体表现形式，再通过资源的载体表现形式关联到资源个体化的单件，呈现出以作品题名为线索的自上而下树形结构；(2)在作品层内部，以作品自身为主体，包括个人和机构在内的责任者与作品之间具有多种责任关系，同时作品表达了相应的主题；(3)在书目记录层内部，作品的载体表现和内容表达被特别分开描述，使得作品的揭示不再根据载体表现区分不同的编目模板；(4)馆藏实体层由包含作品层和书目记录层某一具体属性的单件所构成的集合。

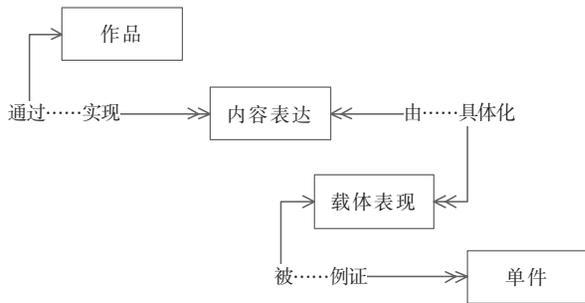


图1 FRBR第一组实体结构

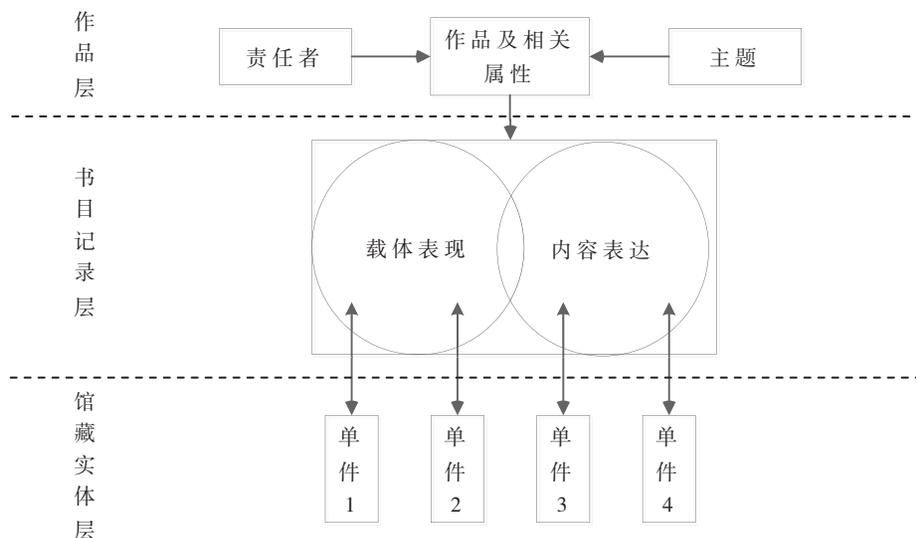


图2 科技文献资源揭示的WBC模型

3 框架元素

在本文提出的科技文献资源揭示模型的基础上，给出科技文献资源描述的元素，WBC模型的每层元素均可分为3种类型，即实体元素集、属性元素集和关系元素集。属性元素集反映一个实体某方面或者某部分的属性特征；实体元素集即所有实体的集合；上述两个元素集的内部和相互之间均存在关联关系，这些关系通过关系元素集来描述。

3.1 规范聚类层

规范聚类层包含3个元素集，分别是责任者、作品及自身属性和主题。

(1) 资源的责任者

科技文献资源的产生、出版、传播、收藏等各个环节的实现主体是个人或团体，对资源的各种责任方式（作者、编者、译者、主办机构、出版机构等）需要进行揭示，以资源的责任者为主体来汇聚所有相关资源。对于元素的选取考虑下列因素：①实现数据关联的前提是具有唯一标识号以及所关联数据的唯一标识号；②确定核心元素来实现责任者的区分性及识别性（如个人名称、团体名称）；③确定非核心元素来保证著录的其他原则和功能的实现（如所属机构），此部分可以根据揭示关系的需求及资源特点进行可选设置。部分描述元素见图3。

(2) 作品及相关属性

该元素集实质为作品和作品的内容表达，可将书目记录各个元素都视为作品，分别建立规范记录，通过关联说明及关系说明元素来帮助规范记录反映作品之间的关系，例如连续出版物继承与被继承的关系。资源元素集是实现数据关联/层级展示的关键元素集，主要实现资源与资源间的关系及资源本身内部的关系，资源的规范记录要将属性细分到最小颗粒度，所以对不同属性的题名名称建立不同的规范记录。元素的选取考虑下列因素：①具有能够识别规范记录的唯一标识号；②与不同属性题名的相关元素，能够更完整、更充分地描述题名实体；③资源规范记录间

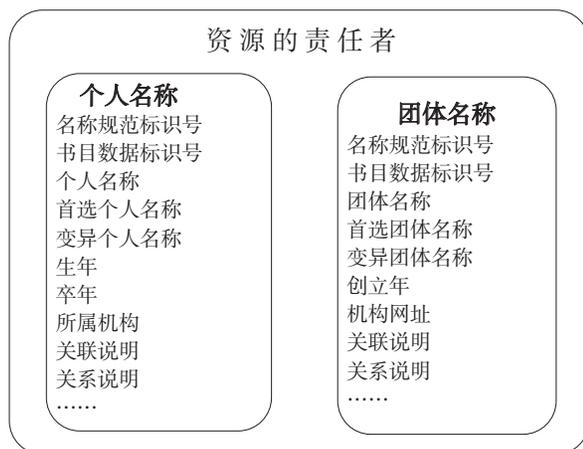


图3 资源责任者描述元素集

关联及资源规范与书目记录的关联元素。部分描述元素见图4。

(3) 主题

主题元素集是指对资源内容层面的揭示，在揭示时需充分利用各个机构的分类体系成果，从学科角度对科技文献资源进行揭示。所选取的元素包括记录ID号、关键词、分类法等，这里不再详细列出。

3.2 书目记录层

本框架的书目记录层具有以下特点。

(1) 针对物理实体进行资源描述的，比如不同机构收录相同报告的出版物要分别进行描述。

(2) 充分吸收AACR2中文献类型特征，选用专指属性的元素进行描述，为实现资源识别、聚类提供必要的基础信息。

(3) 本层次的书目数据为规范层次及馆藏层次的数据提供了基本信息，丰富而准确的著录数据为数据关联打下良好基础。

(4) 元素可以根据资源特点的变化进行适当的增减。

(5) 每条书目记录中设计有关联表现形式及关系说明字段，将作品表现形式间的关系进行揭示。

书目记录层的部分描述元素如图5所示。

3.3 馆藏实体层

馆藏实体层是本框架的最底层，也是直观面向用户服务的资源揭示层。有的馆藏单件实物，其原始内容与其他版本并无区别，只是由于来源渠

道不同,或属非正常出版的实物,如复印本、预印本等,在馆藏形态上形成了一种新的表现形式,这也可以在单件上构成一件新的版本。判断这类版本时需要编目人员核对内容、页码等信息是否完全一致。这里元素集主要包括:馆藏标识号、单件标识号、载体形式、关联单件、关系说明等。

4 元数据格式记录的关联关系

基于上述元数据基本框架,对书目记录间关系的揭示实为对元数据格式记录间关系的挖掘。因此,通过分析元数据格式记录及字段间的关系并进行相应扩展后,可以得到更丰富的书目关系。UNIMARC格式中书目关系为垂直关系、平行关系和年代关系3种;后来对书目记录间的关系做了更细致的揭示,将书目关系分为等同关系、描述关系、继承关系、整体-部分关系、连续关系、同伴关系和共有特性。用于描述网络资源的Dublin Core使用关联元素描述关系,元数据方案包括整体-部分关系、版本关系、格式区别、参引关系等6种关系。MARC、DC等元数

据旨在揭示图书馆传统书目资源记录间的关系。FRBR实体-关系模型将书目资源分为作品、内容表达、载体表现和单件,将书目关系分为高层实体关系、主题关系、责任关系、内容关系、其他关系,并延伸到责任者-资源、主题-资源等。此种关系表现在字段间的关系,为书目关系提供新的思想。

本框架主要以责任方、资源、主题为3条主线将庞大的科技文献资源形成有层级关系的结构化体系。从书目记录分别抽取责任方(个人名称、团体名称)、资源(题名、期刊题名、会议题名等)、主题(关键词、分类号)的对应元素信息,不同属性的元素分别建立规范记录,每条规范记录挂接相应的书目记录。

在关联实现中,将责任方、主题规范记录中的书目ID号与资源的规范记录中的书目ID号进行匹配,将匹配上的书目ID号与书目记录层中书目ID号以及馆藏实体层中的书目ID号进行挂接;资源与资源间的关系由资源规范记录中的关联字段来实现。在实现规范ID号与书目ID号间的关联时,聚类规范层中,不同规范ID号对应相同书目ID号时,表示责任方与资源的关系、主题与资源的关系、资源与资源的关系;规范ID号对应不同书目ID号时,表示责任方、主题对应多个资源,资源对应多个表现形式。在实现规范汇聚层中各个元素集间的关联时,责任方与资源间的关系是指责任方创作了资源,实现了资源的内容表达;主题与资源间的关系是资源以概念、物体、事件、地点为主题;资源内部的关系包括整体与部分关系、连续关系、附属关系、包含与被包含关系、衍生关系等;责任方与书目记录的关系包含收藏关系、出版关系、贡献关系。由此数据构建的关联如图6所示。

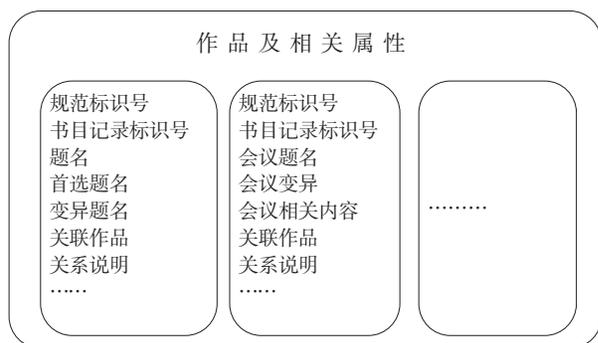


图4 作品及相关属性描述元素集



图5 书目记录层描述元素集

5 应用实例

本文选取 Society of Photo-optical Instrumentation Engineers (SPIE) 相关的科技文献资源为例,初步给出与CNMARC的映射关系以及关联数据的展示^[10]。基本框架元素与CNMARC/USMARC

的映射关系如表1所示。

文献资源物理实体进行信息描述著录，形成书目记录，规范汇聚层的主题、作品、责任方3个元素集的基本信息从书目记录进行抽取，主题元素集抽取的元素有001、6XX主题分析块的相关字段，作品元素集抽取的元素有001、200、225、102、711（12）、4XX款目连接块的相关字段，责任方元素集抽取的元素有7XX知识责任块的相关字段，通过规范ID号与书目记录ID号进行关联以及元素集中的关联关系元素，将与SPIE相关的所有科技文献资源揭示出来，同时也将相互间的关系呈现出来（图7）。左边框中original formats/language/date属于作品的内容表达，online format是作品的表现形式，contributor是作

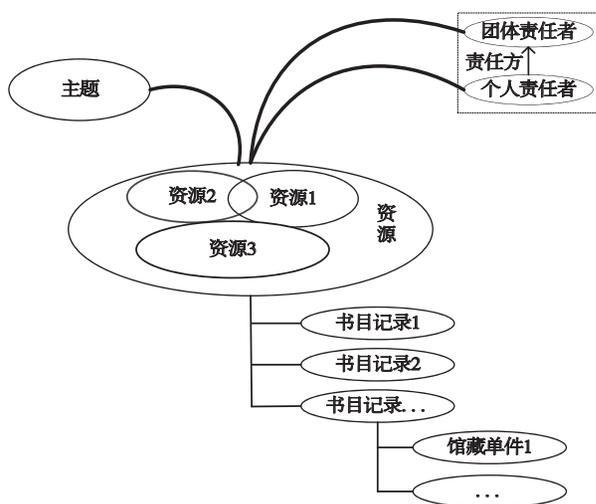


图6 关联关系结构图

品的创作者，subject是主题。右边书目数据的层级关系是作品内部之间的关系，包括等同关系、整体与部分关系、连续关系等，例如SPIE主办或是出版会议，每个会议包含了历届会议，每届会议对应一个或者多个会议录，每个会议录对应一个或者多个单件。

用户在检索框中输入任何关键词，都可以将其相关的信息显示出来，供用户选择，同时也能帮助用户发现未知信息。

SPIE conference proceeding MARC格式元数据样例如下。

```

001 42A0162723
010 $a0-8194-0743-7
035 $a91-641811
100 $a20090602d1991 km y0engy01031213ba
1010 $aeng
102 $aUS
105 $aa z 101yy
106 $ar
2001 $aVisual communications and image process-
ing '91: Image processing$hp art 1$eVisual communications
and image processing conference, 11-13 November, 1991,
Boston, Massachusetts$fedited by Kou-Hu Tzou, Toshio
Koga
225 $aProceedings of SPIE-The International
Society for Optical Engineering$vv ol. 1606, part 1
690 $aTP391.4$v4
71102 $aSPIE-The International Society for Optical
Engineering$4sponsor
    
```

表1 基本框架元素与CNMARC/USMARC的映射关系

| 元素 | CNMARC | 元素 | CNMARC | 元素 | CNMARC |
|--------|--------|-------|------------|-------|--------|
| 正题名 | 200\$a | 期刊开始年 | 100\$a | 出版地 | 210\$a |
| 分辑号 | 200\$h | 期刊终止年 | 100\$a | 出版者 | 210\$c |
| 分辑名 | 200\$i | 会议题名 | 711(12)\$a | 出版年 | 210\$d |
| 丛编题名 | 225\$a | 会议时间 | 711(12)\$f | 附件 | 215\$e |
| 丛编ISSN | 225\$x | 会议届次 | 711(12)\$d | 关键词 | 610\$a |
| 丛编分辑号 | 225\$v | 会议地点 | 711(12)\$e | 中图分类号 | 690\$a |
| 期刊题名 | 200\$a | 个人责任者 | 701\$a | 杜威分类号 | 676\$a |
| 先前题名 | 430\$a | 团体责任者 | 711(02)\$a | 馆藏机构 | 905\$a |
| 后续题名 | 440\$a | 正文语种 | 102\$a | 馆藏号 | 905\$h |
| 期刊ISSN | 011\$a | ISBN | 010\$a | 馆藏媒介 | 905\$n |
| 先前ISSN | 430\$l | ISSN | 011\$a | | |
| 后续ISSN | 440\$l | 报告号 | 096\$a | | |

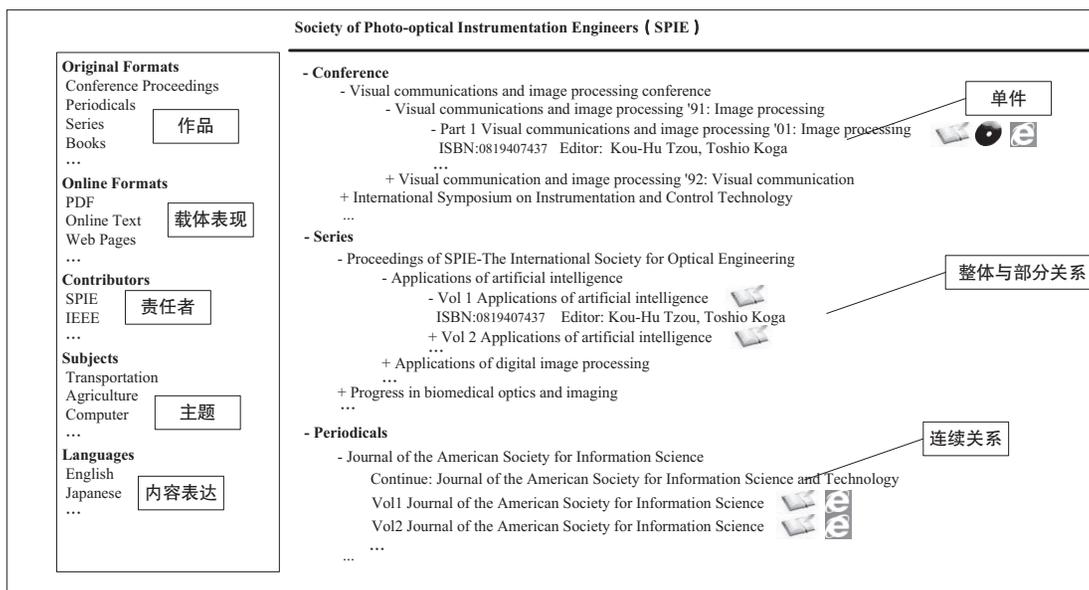


图7 检索机构名称为SPIE时所显示的结果

7112 \$aVisual communication and image process-
ing conference\$f19911111-13\$eBoston, MA(US)
905 \$aCN111001\$cEM\$dTP391.4\$h9201108
905 \$aCN111001\$cEM\$dTP391.4\$hIEL9418\$non
line
SPIE Periodicals MARC格式元数据样例如
下:
001 12A0070993
011 \$a0002-8231
040 \$aAISJB6
092 \$d375B0002
100 \$a20080123b19702000km y0engy-
01031234ba
1010 \$aeng
102 \$aUS
106 \$ar
110 \$aaza |||0||
2001 \$aJournal of the American Society for
Information Science
2070 \$aV.21(Jan./Feb.1970)-v.51,no.14(Dec.2000)
210 \$aNew York, NY\$cJohn Wiley & Sons
Inc.\$d1970-2000
326 \$a14 issues a year\$b1998-2000
430 \$1011 \$a0096-946X\$12001 \$aAmerican
documentation
440 \$1011 \$a1532-2882\$12001 \$aJournal of the
American Society for Information Science and Technology

531 \$aJ. Am. Soc. Inf. Sci.
6060 \$aInformation science
6060 \$aDocumentation
676 \$a020/.5
680 \$aZ1007
690 \$aG25\$v4
71102 \$aSociety of Photo-optical Instrumentation
Engineers (SPIE)
905 \$aCN111001\$bG-013\$cEP\$dG25\$s4-148
905 \$aCN111001\$bE163EP00003\$cEP\$dG25\$no
nline

6 结论

(1) 目前使用的MARC格式元数据框架已经不能满足数字图书馆时代跨载体表现形式、跨资源类型的科技文献资源描述需求, 国内外很多数字图书馆开始关注科技文献的知识组织。然而, 不论是FRBR/FRAD框架还是RDA编目规则, 虽能够适应这种背景, 却也有着自身局限性。比如: FRBR/FRAD框架仅仅着眼于概念框架本身, 并无具体的编目规则与其对应; RDA编目规则则由于其从根本上挑战并改变了传统的资源揭示方法, 能够实现多角度的数据关联、更深层次的资源揭示, 但由于与传统编目规则有根本性的改变, 我国对RDA还处于观望状态。

(2) 本文基于FRBR概念模型的思想, 结合目前MARC元数据框架的特点, 将两者最大限度地有机结合, 设计出一种向RDA规则过渡的元数据框架, 打破了依照文献类型逐章描述各自著录规则的体系, 将所有的文献类型元数据框架划一, 形成通用基本框架, 从基本框架提取所需揭示的基本信息形成著录单元细分的元素集, 这使得不同载体不同文献类型的著录框架具有高度的连贯性和一致性。从可操作性上来看, 本框架可以继续沿用现有系统, 现有的编目人员不需进行系统的培训即可操作, 通过对框架的设计机制让系统处理关联关系, 为将来选择RDA做好了准备, 奠定了基础。另外, 通过本框架还可对科技文献资源的内部实体进行再组织, 在揭示中可实现更好的可视化效果。

(3) 由SPIE的例子可以看出, 本框架能够按照多维度(如文献类型、载体、语种、主题、责任者等)进行揭示, 每种维度中也展示了书目记录间的关系(如整体与部分关系、连续关系等), 给用户读者展示出系统全面的检索结果, 便于其挖掘其他重要文献。图书馆可以利用本框架体系可自行组织本馆特色资源的功能, 呈现给用户读者全面系统的资源界面。

(4) 本框架能够适应互联网服务的需求, 谷歌、百度等互联网工具可以提供检索, 但存在检全查准的问题, 而图书馆可以利用自身书目记录的优势, 通过数据关联使用户读者挖掘出其他需要的文献。同时, 可以与出版发行方的元数据进行对接, 使用户读者了解出版发行的情况。

(5) 本文提出的基本框架有其局限性。从记录本身来看需要数据信息项足够丰富、完整, 并且各个检索点具有规范文档控制, 投入一定的人

员对记录间的关系进行说明; 从文献资源类型上看, 图书文献类型关系揭示需要完整的规范记录支撑。本框架适用于期望历史编目数据能够可视化关系展示的情况以及本馆还未采用RDA规则但希望能够可视化揭示关系的情况。

参考文献

- [1] 辜寄蓉, 苗放. “数字成都”综合信息资源管理平台框架设计[C]//中国测绘学会. 第四届海峡两岸GIS发展研讨会暨中国GIS协会第十届年会论文集. 2006:519-523.
- [2] 金更达, 何嘉荪. 数字档案馆模式探讨——基于元数据的电子文件集成管理与服务研究之二[J]. 档案学通讯, 2005(5):54-58.
- [3] AustLit. The Australian Literature Resource Home Page [EB/OL]. [2012-07-10]. <http://www.austlit.edu.au/>.
- [4] Timothy J Dickey. FRBRization of a Library Catalog: Better collocation of Records, Leading to Enhanced Search, Retrieval, and Display[J]. Information Technology and Libraries, 2008, 27(1):23-32.
- [5] 谢琴芳. FRBR用户任务对未来图书馆目录影响[C]. 中国图书馆学会2005年年会, 2005:56-59.
- [6] 鲍国强. FRBR基本模式在古旧地图编目及数字化信息检索中的应用[C]//第一届全国文献编目工作研讨会论文集. 2006:39-49.
- [7] 李丹. RDA的目标与原则[EB/OL]. [2014-06-03]. http://www.nlc.gov.cn/newtsgj/gtqk/tyck/2009nzml/111112/zl2/201012/t20101201_23723.htm.
- [8] 胡小菁. RDA实施前景分析[J]. 数字图书馆论坛, 2012(9):38-43.
- [9] 傅西平, 孙更新. RDA的普及难度及发展趋势——以斯坦福大学图书馆RDA测试为例[J]. 情报杂志, 2013, 32(8):132-135.
- [10] 胡小菁. CNMARC的RDA更新[J]. 数字图书馆论坛, 2013(7):2-7.